

וויזואליזציה של מידע

הקשר בין הורים לסיכוי של הצאצא להפוך לכלב
נחיה

מגישים:

עדי פרלוב, 200477842

רעות מזרחי, 305160079

Overview:

כלב נחייה - כלב שנועד לסייע לאדם עם לקות. הסוג הנפוץ ביותר של כלבי נחייה עוזר לניידותו של אדם לקוי ראייה, ובפרט עיוור.

כלבי נחייה נבחרים, על פי רוב, מגזעי כלבים חזקים ושקטים כמו לברדור רטריבר, גולדן רטריבר, פודל, בורדר קולי, לברדודל ורועה גרמני או מבן כלאיים בין הסוגים. בישראל נעשה שימוש בעיקר בגולדן רטריבר ובלברדור רטריבר.

כיום בארץ יש את ה"מרכז הישראלי לכלבי נחיה" אשר מייצר בכל שנה 35 צמדים חדשים של אדם עם לקות ראייה וכלב נחיה.

אימון הכלבים, הצמדת האדם עם לקות הראייה והשילוב ביניהם הוא תהליך מורכב שנמשך כשנתיים ועלותו גבוהה.

הכלבים הם צאצאים של כלבי לברדור וכלבי רטריבר מובחרים. בגיל 8 שבועות הגורים נמסרים למשפחות אומנות למשך שנה, במהלך השנה הזו הגורים נחשפים למגורים ציבוריים, תחבורה ציבורים וכו' על מנת שיתרגלו למצבים בהם יצטרכו להתמודד בעתיד.

בסוף השנה הכלבים חוזרים למרכז ועוברים תהליך של הערכה ומיון, כלבים בעלי תכונות מתאימות אשר עברו את תהליך המיון (מבחנים) מתחילים תוכנית אימונים בת חצי שנה.

בתהליך המיון הכלבים עוברים מבחנים רבים אשר בודקים את מידת ההתאמה של הכלב להפוך לכלב נחיה ולסייע בהצלחה את תהליך האימון, מבחנים לדוגמא : האם הם מפחדים מזרים, או חוששים שיגעו בהם ועוד...

Data Base:

בפרויקט השתמשנו ב- Data base אשר נלקח מאתר הקורס "ויזואליזציה של מידע" כל הנתונים נאספו על ידי "המרכז הארצי לכלבי נחיה" ורוכזו על ידי ד"ר אנה זמנסקי.

קבצי ה- Data Base מצורפים כחלק מקבצי ההגשה: **dogs_status.csv, test.csv**.

ה- Data base מחולק לשני חלקים, האחד כולל את ה- attributes הבאים:
Mother, Father, Status. הסטטוס יכול לקבל את אחד מהפרמטרים הבאים:

- Guiding
- Specialneeds
- Unsuitable

החלק השני כולל מידע אודות הטטטים שעברו הכלבים, והוא כולל Mother name, Father name ועוד 51 טטטים שונים.

עבור כל טסט ישנו פרמטר בוליאני (reason) המציין האם בעקבות טסט זה הכלב נכשל (יכולים להיות יותר מטסט אחד).

חשוב לציין שבנתונים היו כלבים אשר לא עברו את המיונים כלומר קיבלו את הסטטוס Unsuitable או Special needs אך לא נכשלו באף טסט. לכן הוספנו סיבת כישלון נוספת בשם: "Other".

Task 1:

המטרה העיקרית בפרויקט היא להבין לפי ויזואליזציה של ה data base הנתון, האם קיים קשר בין זוג הורים מסוים לבין סיכויי ההצלחה של הצאצאים שלהם להפוך לכלבי נחיה.

לדוגמא: אם ישנם זוג הורים, מוקה וגיבור שיש להם שישה צאצאים וכולם בסטטוס "Special Needs" או "Unsuitable" ניתן להסיק מכך שהגנים של הכלבים משפיעים על הצלחתם במיונים לכלבי נחיה וכי הסיכויים של הצאצאים של מוקה וגיבור להיות כלבי נחיה הינם נמוכים.

Task 2:

ב- Task זה אנו בוחנים את הקשר בין זוג הורים מסוים לטסטים שצאצאיהם נכשלו בהם. המטרה בוויזואליזציה זו להבין האם הגנטיקה שהורים מעבירים לילדיהם משפיעה על מידת ההצלחה בטסטים מסוימים.

כלומר, האם צאצאים אשר קיבלו את הסטטוס Unsuitable או Special Needs נכשלו במספר גבוה של טסטים משותפים ניתן להסיק שקיים קשר בין זוג הורים לכישלון בטסט כזה או אחר.

ניתוח נתונים:

עבור הוויזואליזציה של "הקשר בין זוג הורים לטסטים שצאצאיהם נכשלו" ב- Task השני היינו צריכים לשלוח את המידע הרלוונטי מה- Data Base רק עבור זוג ההורים שנבחר ורק עבור ילדים שנמצאים תחת הסטטוס Unsuitable או Specialneeds.

לכן, המידע עבר סינון ע"פ פרמטרים אלו.

למה זה מעניין?

ראשית, נציין שעלות אימון כלב נחיה יקרה מאוד, היא דורשת זמן ומשאבים רבים. בנוסף, ידוע שלא כל כלב המתחיל את התהליך כשיר לסיים אותו מסיבות שונות. ולכן, אם הייתה את היכולת להעריך האם כלב יסיים את ההכשרה, היה ניתן לחסוך כסף רב וזמן יקר. על ידי התבוננות בוויזואליזציה, ניתן לראות האם סיכויי ההצלחה של כלבים לסיים את תהליך ההכשרה ולהפוך לבסוף לכלבי נחיה יכולים להיות מושפעים מהגנים שההורים שלהם מעבירים להם, וגם, האם צאצא מסוים עתיד להיכשל בטסט כזה או אחר רק משום שנולד לזוג הורים. זו נקודה שיכולה לעניין את המרכז הארצי לכלבי נחיה, ובכלל כל גוף או מוסד המכשיר כלבי נחיה. באמצעות הוויזואליזציה הם יוכלו להעריך אילו זיווגים נותנים את הסיכוי הטוב ביותר לצאצא, ואילו זיווגים נועדו לכישלון מראש. וכך למעשה, להביא לחסכון רב גם בכסף וגם בזמן ובמשאבים הרבים הנדרשים להכשרת כלב נחיה.

Patterns

בוויזואליזציה זו נתמקד בשתי תבניות עיקריות:

- תבנית ראשונה: לזוג הורים יש מספר ילדים וכולם קיבלו את אותו הסטטוס. כלומר, נבחר לדוגמא זוג הורים Silver ו-Willow, אם נראה שכל ילדיהם קיבלו את הסטטוס "Guiding" אז נדע שקיים סיכוי גבוה שצאצא שלהם יהפוך לכלב נחיה, ולכן כדאי לזווג אותם על מנת להגדיל את הסיכוי שהצאצא יסיים את תהליך ההכשרה וכך לחסוך בכסף ובזמן. באותו אופן גם בכיוון השלילי, אם לזוג הורים כל ילדיהם קיבלו את הסטטוס "Unsuitable" אז נדע שלא כדאי לזווג ביניהם מכיוון שקיימים סיכויים גבוהים שצאצאיהם לא יסיימו את תהליך ההכשרה ועשויים להיכשל בתהליך המיון.
- תבנית שנייה: טסטים כושלים המשותפים למרבית הצאצאים של זוג הורים. כלומר, אם נראה שישנו זוג הורים שכל הצאצאים שלו נכשלו באותם הטסטים נדע שקיימים סיכויים גבוהים שצאצא שלהם יכשל באותם הטסטים. לדוגמא אם נמצא שכל הצאצאים של Silver ו-Willow נכשלו בטסט "חרדת נטישה" נדע שקיים קשר בין הגנטיקה של ההורים לכישלון בטסט זה.

Visualization - Parallel Sets

על מנת להציג את המידע בחרנו להשתמש ב-Parallel Sets, הסיבה לכך היא שהוויזואליזציה הזו מראה באופן רחבי את מידת ההצלחה של הכלבים כך שהם מסודרים לפי ההורים ולכן זה משקף בצורה הטובה ביותר האם יש קשר ישיר בין ההורים לבין מידת ההצלחה של הצאצאים שלהם במבחנים שעוברים על מנת להיות כלבי נחיה.

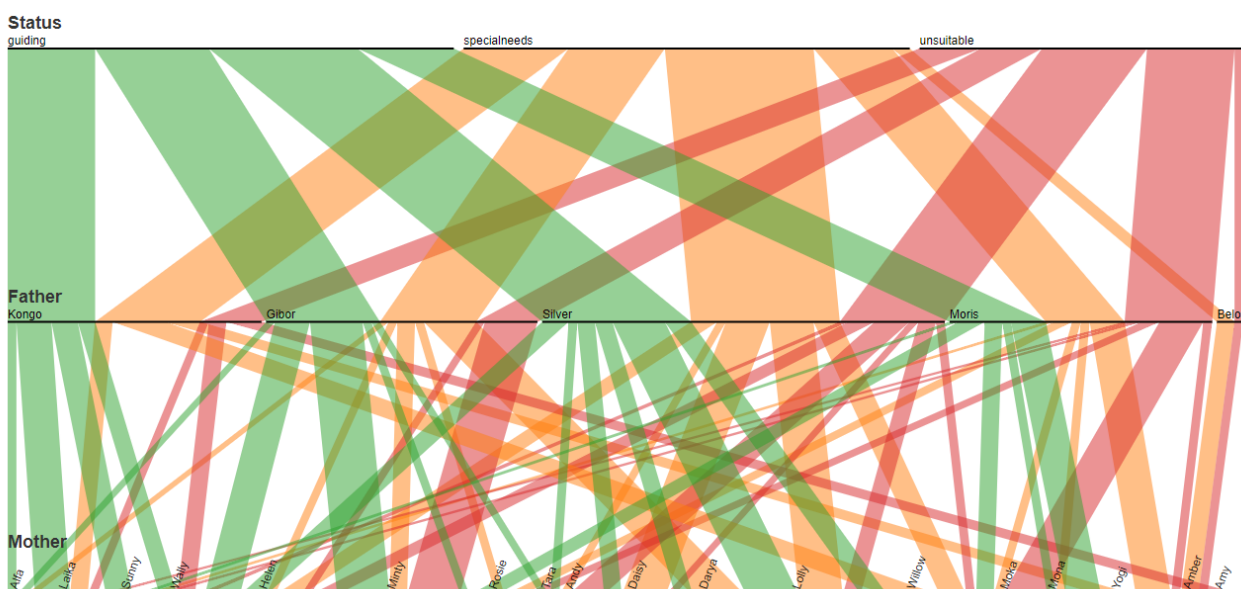
Parallel Sets מוצג כך שכל שורה מתאימה לממד אשר הערכים שלו מיוצגים בכל קו בשורה הזו. רוחב כל שורה ונתיב הזרימה הנובע ממנה נקבע על ידי החלק היחסי. כל נתיב זרימה הוא מובדל בצבע ייחודי על מנת להציג ולהשוות בין הקטגוריות השונות.

בפרויקט שלנו קיימים שלושה ממדים:

- Status
- Father
- Mother

בחרנו את הממדים הללו לפי הנתונים הקיימים בData base, את סדר הממדים בחרנו על מנת לשקף בצורה הברורה ביותר את הזרימה בין הסטטוס לבין זוג הורים.

הוויזואליזציה:



צבעים:

בחרנו את הצבעים בצורה כזאת שיהיה טבעי לקורא להבין איזה סטטוס מסווג כ-"טוב" ואיזה סטטוס מסווג כ-"לא טוב".

עבור הסטטוס "Guiding" בחרנו את הצבע ירוק מאחר וירוק מהווה קונוטציה של משהו טוב, שעבר בהצלחה.

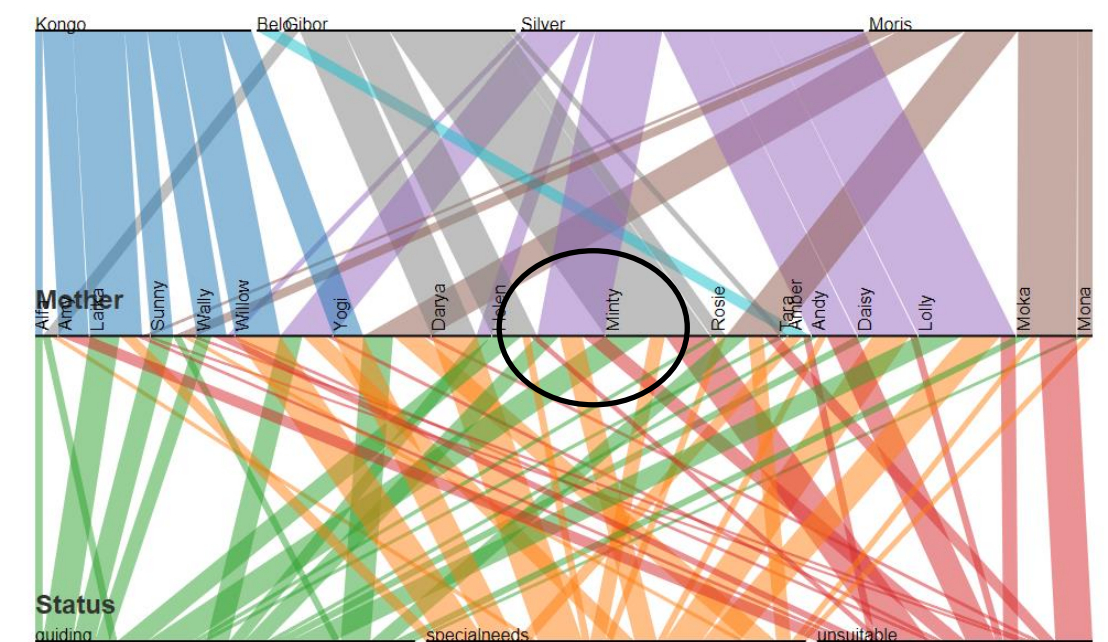
עבור הסטטוס "Unsuitable" בחרנו את הצבע אדום מאחר והצבע הזה מהווה קונוטציה של משהו לא טוב, של כישלון.

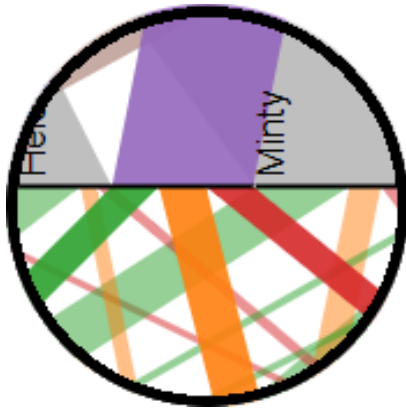
לכן (בהתאמה) בחרנו עבור הסטטוס "Spacial Needs" בחרנו את הצבע כתום, כי הוא ה"אמצע" של הצבעים אדום וירוק ולכן גם משקף את הסטטוס, כלומר "Special Needs" לא מביע כישלון או הצלחה, זה מצב ביניים ולכן הצבע גם מהווה צבע ביניים בין אדום לירוק.

ממדים:

בחרנו בסדר הממדים הנ"ל מכיוון שבסדר הפוך נוצרת בעיה: מכיוון שעבור כל אבא יש יותר מאמא אחת שאליה הוא מחובר, חייבים להשתמש ב-5 צבעים, 1 עבור כל זוג אבא-אמא.

דוגמא לוויזואליזציה עם סדר ממדים שונה:





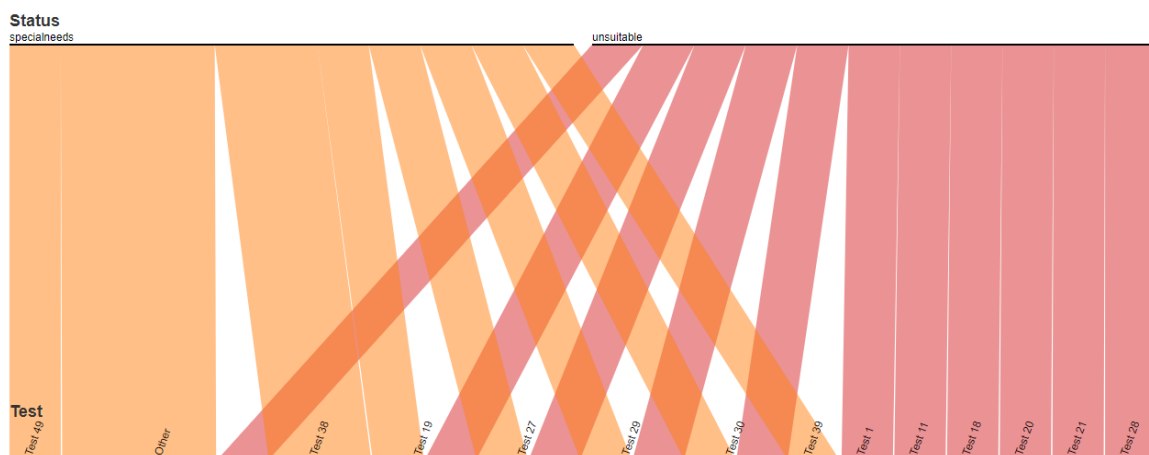
ניתן לראות שבין הכלב silver לכלבה minty יש יותר מסטטוס אחד ולכן הצבע של הזוג (סגול) יתפצל לשלושה צבעים (ירוק, כתום ואדום) ע"פ הסטטוסים המתאימים.
וממצב זה רצינו להימנע.

את ה- use cases חילקנו לשתי וויזואליזציות:

- תצוגה של הצאצאים של כל זוג הורים והסטטוס שלהם.
- עבור כל זוג הורים ניתן לעבור לוויזואליזציה שונה בה ניתן לראות באילו טסטים כל הצאצאים שלהם נכשלו. לוויזואליזציה זו ניתן לעבור בקלות ע"י לחיצה על זוג הורים כאשר יש להם ילדים בעלי סטטוס "Unsuitable" או "Speacialneeds".

חילקנו את הוויזואליזציה לשניים מאחר ולא ניתן היה לאחד את שני ה- Tasks, זאת מאחר שבשביל לאחד הכל הייתה אופציה להוסיף ממד בשם "טסטים" ועבור כל כלב היינו משייכים את הטסט בו הוא נכשל, אך נתקלנו בבעיה כאשר כלב נכשל ביותר מטסט אחד, במצב כזה לא ניתן לייצג זאת. מאחר ובמידה וכלב נכשל בכמה טסטים זה יחשב כמספר כלבים שונים. כלומר ב data base עבור כל כלב ניתן לכתוב ערך אחד בטסטים בו הוא נכשל, לכן כדי להציג מספר טסטים שבו נכשל הכלב נצטרך להוסיף שורה עבור כל טסט בו הכלב נכשל ולכן זה יצור מצב שבו נוספו כלבים שלא קיימים. לכן פיצלנו את הוויזואליזציה עבור task2.

הוויזואליזציה עבור Task 2:



ויזואליזציה זו מציגה את כל הטסטים הכוללים עבור כל הכלבים תחת הסטטוסים "specialneeds" ו-"unsuitable". ברור שאין כאן את הסטטוס "guiding" מכיוון שכלבים עם סטטוס זה לא נכשלו אף לא בטסט אחד.

ניתן לראות בדוגמא לא מעט טסטים כושלים המשותפים ליותר מכלב אחד. מכך ניתן להסיק שהצאצאים עבור אותו זוג הורים שמוצג בתמונה מתקשים יותר בטסטים הספציפיים האלה.

צבעים:

הצבעים בוויזואליזציה זו דומים לקודמת על מנת להקל על המשתמש את ההבנה על מה הוא מסתכל עכשיו. ומכיוון שלהגיע לוויזואליזציה זו צריך ללחוץ על ה-path של זוג הורים רצינו לשמור על קשר ישיר בנוגע לצבעים, לכן "specialneeds" קיבל כתום ו-"unsuitable" קיבל את הצבע האדום.

Prediction

בפרויקט קיימת היכולת לחזות את סיכויי ההצלחה של קומבינציה בין זוג הורים על פי הנתונים הקיימים ב- Data Base כלומר, קיים יישום שמקבל כ- input שם של אב ושם של אם וכ- output מציג את הסיכויים של הכלב להיות כל אחד מהסטטוסים (Guiding, Unsuitable, Special Needs).

Prediction:

Father: Mother:

☐ Guiding: 46.0%

☐ specialneeds: 38.0%

☐ unsuitable: 15.0%

כאן ניתן לראות שעבור כל הצאצאים של הכלב Silver והכלבה Lolly ישנם 46% לסיים את ההכשרה והמיון ולהפוך לכלב נחיה, 38% לא לעבור את המיון לכלב נחיה אבל להיות מספיק טוב להימסר לצרכים מיוחדים וכ-15% לא להתאים לאף מסגרת.

אם מנסים לחזות זוג כלבים שאין להם צאצאים מקבלים הודעה שלא קיים מידע להציג.

Prediction:

Father: Mother:

There is no data to show for **Silver** and **Mona**.

☐ Guiding:

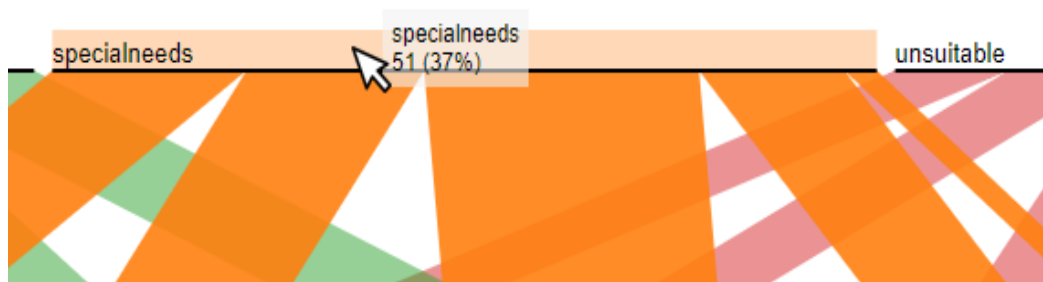
☐ specialneeds:

☐ unsuitable:

How does it look?:

כבר הצגנו קודם לכן את שתי הוויזואליזציות , ולכן כאן נפרט על הפיצ'רים שהכנסנו על מנת להקל את השימוש ולהפוך את הוויזואליזציה לברורה ומובנת יותר ואת חוויית המשתמש לקלה יותר.

Tooltips and highlight

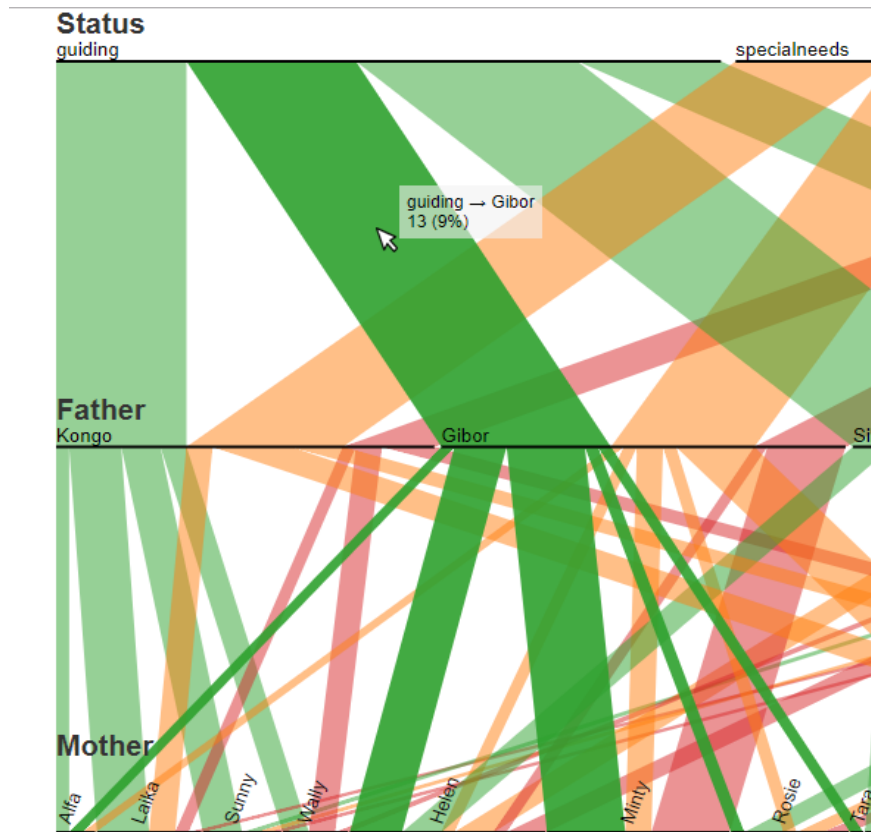


כאשר עומדים עם העכבר על ה- attribute של ממד מסוים, אז מופיע חלון tooltip עם השם של אותו attribute ועם מספר האלמנטים ב- Data Base עבור ה- attribute ועם האחוז שהם מהווים מהכלל.

למשל כאן ניתן לראות שישנם 51 כלבים בעלי סטטוס "specialneeds" והם מהווים 37% מכלל הכלבים.

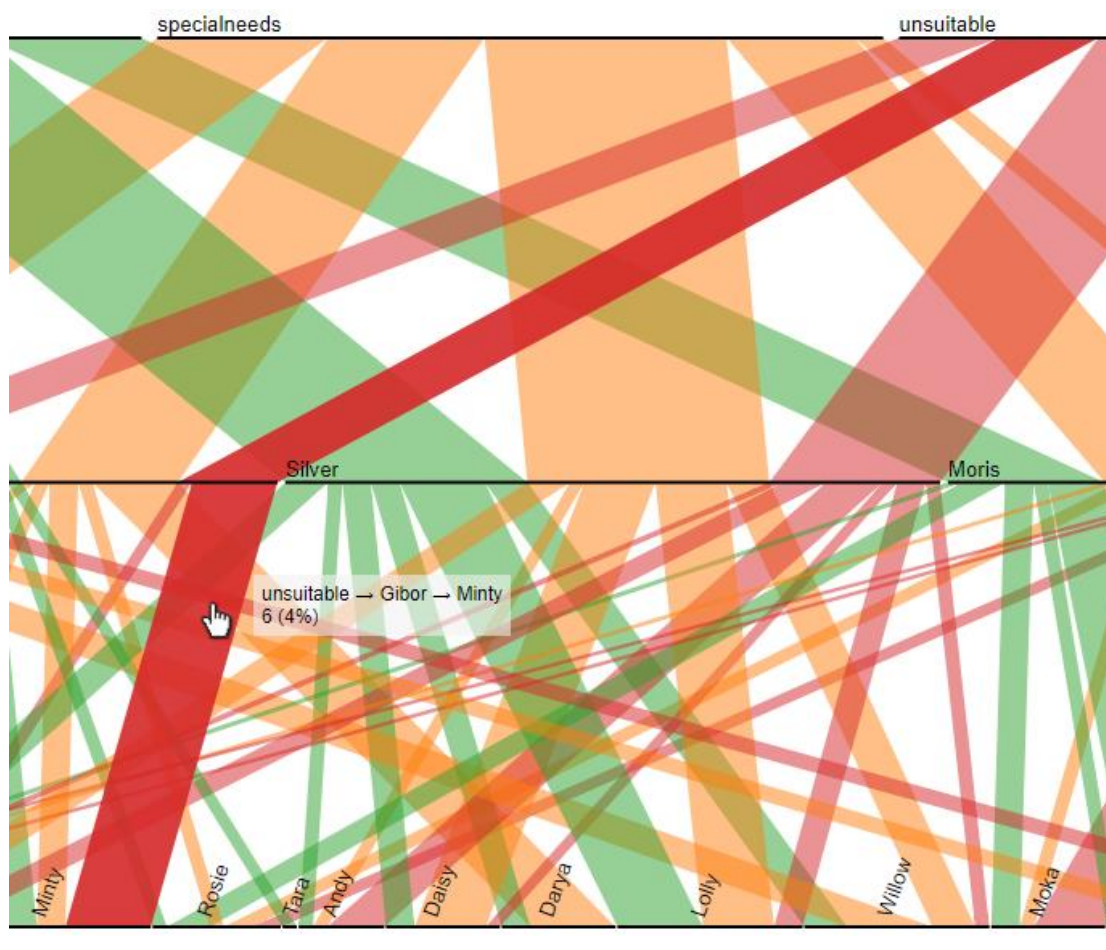
Highlight paths:

כאשר עומדים עם העבר על path מסוים הוא מסמן אותו וגם נותן אינדיקציה לאיזה סטטוס הוא מקושר.



כאן ניתן לראות סימנו את הצאצאים של הכלב Gibor, יש לו 13 צאצאים שמהווים 9% מכלל הכלבים.

לGibor יש צאצאים בעלי סטטוס guiding עם הכלבות Alfa, Helen, Rosie, Tara. עובי ה-path תואם ליחס.



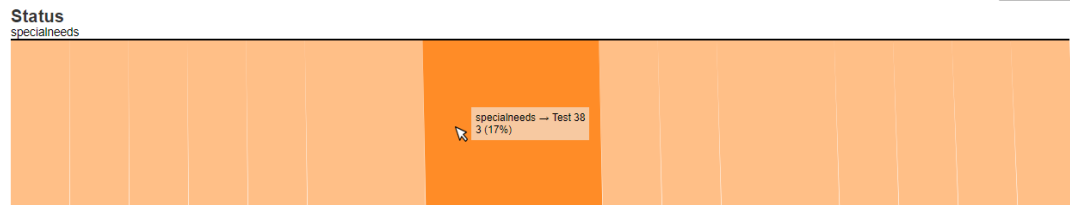
כאן ניתן לראות שישנם 6 כלבים בעלי סטטוס "unsuitable" עבור זוג הכלבים Gibor (לא בתמונה) ו-Minty שמהווים 4% מכלל הכלבים.

ניתן לראות גם שסמן העבר הוא מכיוון שע"י לחיצה עובר לוויזואליזציה עבור Task 2.

הסימן יופיע עבור זוג הורים כאשר הסטטוס הוא "unsuitable" או "speacialneeds" בממד האחרון.

Tooltip in failed tests visualization

חלון ה- tooltip בוויזואליזציה של ה- failed tests מייצג את מספר הכלבים עבור זוג ההורים שנבחר אשר נכשל בטסט הספציפי עליו עומדים עם סמן העכבר.



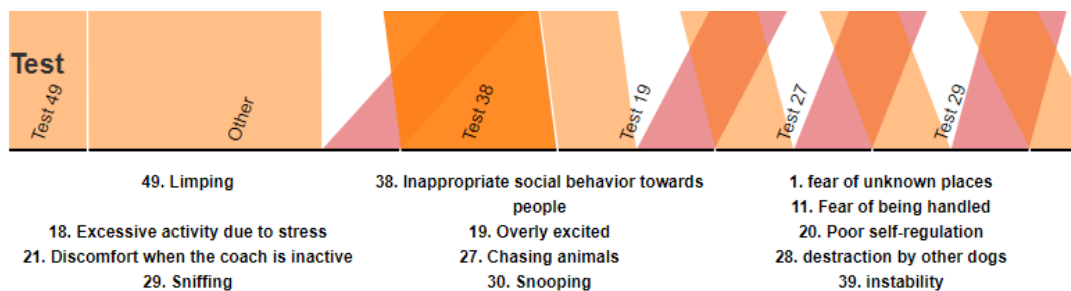
בדוגמא זו ניתן לראות ש-3 כלבים שונים בעלי סטטוס "specialneeds" נכשלו בטסט מספר 31 (מפורט ב- dictionary את השם המלא של הטסט) וכישלון בטסט זה מהווה 17 אחוז מכלל הטסטים שנכשלו.

בנוסף, ניתן לראות שעובי ה- path הוא הגדול ביותר ביחס ל- 17%.

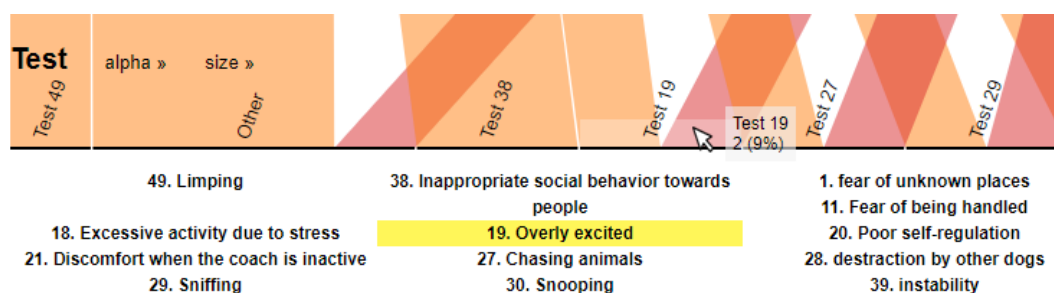
Dictionary

ב-2 task בחרנו לעשות "מקרא" במקום לכתוב את שמות הטסטים בגרף מאחר וישנם שמות מאוד ארוכים ומצד שני יש שמות קצרים ולכן זה לא היה נראה טוב.

על מנת לפתור את הסוגיה הנ"ל, החלטנו לכתוב בתוך הגרף את מספרי הטסטים ומתחת לגרף כתבנו עבור כל מספר טסט את שמו כפי שהופיע במסד הנתונים.



כאשר עוברים עם העכבר על טסט מסוים, הוא נצבע בצבע צהוב על מנת להקל למשתמש את החיפוש של אותו טסט.



מכיוון שוויזואליזציה זו מופיעה באופן דינמי לאחר לחיצה על זוג הורים, הוספנו מידע נוסף שיהיה ברור למשתמש איזה מידע הוא רואה כרגע.

Information:

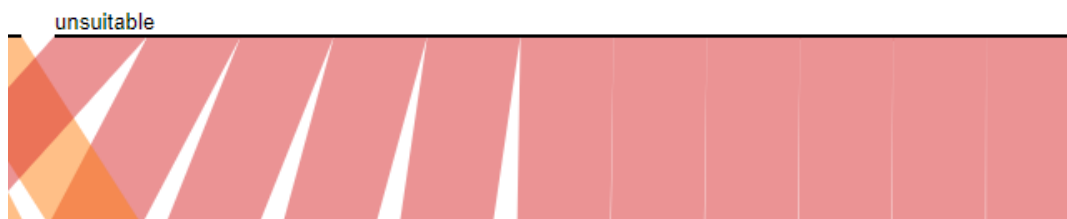
The failed tests of **Willow** and **Silver** offsprings.
The following data refer to 4 different dogs.

הוויזואליזציה מתייחסת לטסטים הכושלים של הצאצאים עבור זוג ההורים Silver ו- Willow. הטסטים הכושלים מתייחסים ל-4 כלבים שונים בעלי סטטוס "specialneeds" ו- "unsuitable" והם צאצאים של Silver ו- Willow.

כפתור חזרה

על מנת להקל את שימוש באתר, הוספנו כפתור חזרה שמחזיר באופן דינמי את המשתמש לוויזואליזציה עבור Task 1.

Go Back

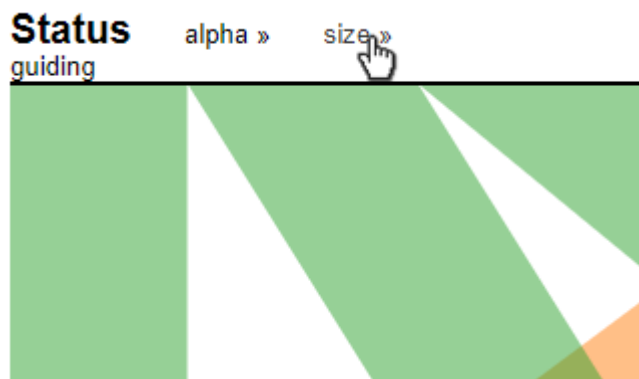


Sorting:

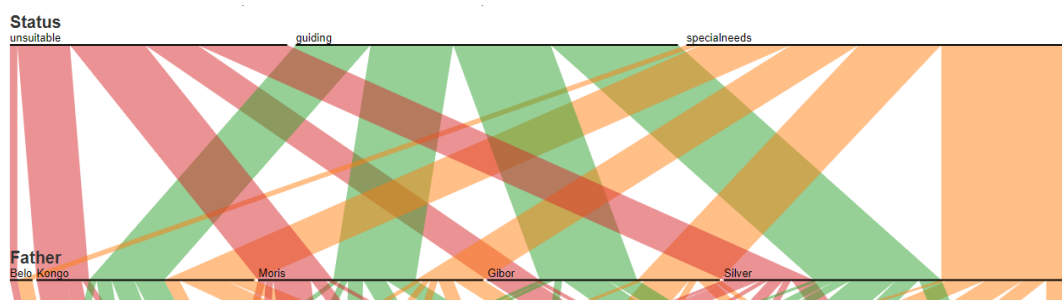
ישנן שתי אפשרויות למיין את הוויזואליזציה:

עבור כל ממד:

- Alpha - ממין על פי סדרת האלף-בית, ניתן למיין מהקטן לגדול ולהיפך.
- Size - ממין על פי גודל, גם במקרה זה ניתן למיין מהקטן לגדול ולהיפך. כלומר אם אנחנו רוצים למיין למשל מהגדול לקטן אזי הסטטוס הנפוץ ביותר יופיע ראשון והסטטוס שמופיע הכי פחות יוצג אחרון.



אפשרויות המיין יופיעו כאשר העכבר עובר מצד ימין לשם הממד.



בדוגמא זו מימד הסטטוס ומימד האבות ממויינים מהקטן לגדול.

Value Of Visualization:

כעת נעריך את הוויזואליזציה בהתאם לטרמינולוגיה שנלמדה בקורס.

יתרונות:

- צבעים – הצבעים ב parallel sets מחולקים לפי הסטטוס נותנים מחדדים את ההבנה של סטטוס טוב ופחות טוב.
- יחס – יש הערכה טובה של היחס כלפי שאר הסטטוסים.
- מימדים – ה- parallel sets מאפשר חלוקה למימדים כך שניתן לעקוב אחרי ה- path של כל זוג הורים עבור כל סטטוס.
- עובי ה- path – העובי של ה- path הוא ביחס ישר לאחוז היחסי של אותו attribute ולכן, קל למשתמש להבין איפה הנתונים עם ה- "בשר" רק ממבט ראשון.
- Tooltip – חלון ה- tooltip עוזר להתמצאות בתוך ה- path.
- Highlighting – עוזר לשים דגש על ה- path הספציפי שהמשתמש רוצה לבחור מבין שאר ה- paths.
- Sorting – כלי המיון עוזר להתמצא ביחס בין הגדלים הן של הסטטוס והן של מספר הכלבים עבור כל הורה.
- ה- parallel-sets מרכז הרבה data תחת ויזואליזציה אחת.

חסרונות:

- עומס במימד האימהות – ישנן 17 אמהות שונות ולכן במימד זה יותר קשה להתמצא.
- מכיוון שיש עומס בממד של האמהות וכל path מקבל את גודלו היחסי אז path קטן יחסית "נבלע" באחרים.
- מימד ה- Test – ב- parallel sets לא היה ניתן לתמוך במימד זה תחת ויזואליזציה אחת, מכיוון שעבור כלב אחד ניתן לתת ערך יחיד, ולכן היה צריך לפצל את הוויזואליזציה.
- Labels – אין הרבה מקום לשמות ארוכים, יוצר עומס מיותר על הוויזואליזציה. מכיוון שיש 51 טסטים היינו צריכים לבצע הוספה של Dictionary.
- הצבעים תלויים במימד הראשון, כפי שהראנו קודם לכן, ולכן סדר המימדים מושפע מכך.